

NPS ARCHIVE
1963
EATON, T.

DISCRIMINATION PROCEDURES,
SMALL SAMPLE PERFORMANCE
THOMAS E. EATON

LIBRARY
U.S. NAVAL POSTGRADUATE SCHOOL
MONTEREY CALIFORNIA

DISCRIMINATION PROCEDURES,
SMALL SAMPLE PERFORMANCE

* * * * *

Thomas E. Eaton, Jr.

DISCRIMINATION PROCEDURES,
SMALL SAMPLE PERFORMANCE

by

Thomas E. Eaton, Jr.

Lieutenant, Supply Corps

United States Navy

Submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE
with major in
MATHEMATICS

United States Naval Postgraduate School
Monterey, California

1 9 6 3

DISCRIMINATION PROCEDURES,
SMALL SAMPLE PERFORMANCE

by

Thomas E. Eaton, Jr.

This work is accepted as fulfilling
the thesis requirements for the degree of
MASTER OF SCIENCE
with major in
MATHEMATICS
from the
United States Naval Postgraduate School

ABSTRACT

The general two population discrimination problem is discussed briefly under various situations. Discrimination procedures using the linear discriminant function and a nonparametric procedure due to J. L. Hodges and E. Fix which classifies a random variable to a population on the basis of assigning it to the population which has the nearest observation to an observed value of the random variable are discussed and compared by computing the probabilities of misclassification for both procedures when the two populations are normal with equal covariance matrices. Probabilities of misclassification are computed for the nonparametric discriminator and the linear discriminant function for two small sample sizes for the case when the two populations being discriminated are exponential. In this latter case, both discrimination procedures are shown to give high probabilities of misclassification for certain values of the parameters of the distribution being discriminated. Regions are given in terms of the parameters of the two exponential distributions where one of the probabilities of error is greater than 0.5. A more complete investigation for larger sample sizes is recommended for the linear discriminant function and the nonparametric procedure discussed in this paper for the case when the two populations being discriminated are exponential.

TABLE OF CONTENTS

Section	Title	Page
I.	Introduction	1
II.	Performance of the Linear Discriminant Function and a Class of Nonparametric Discriminators When the Two Populations Being Discriminated Have Normal Distributions With Equal Covariance Matrices	5
III.	Performance of the Linear Discriminant Function and a Class of Nonparametric Discriminators When the Two Populations Being Discriminated Have Exponential Distributions	19
IV.	Summary and Conclusions	26
V.	Bibliography	29

SECTION I

INTRODUCTION

The two population discrimination problem may be summarized as follows: given a random variable Z distributed over some p -dimensional space according to a distribution F , or according to a distribution G , determine on the basis of an observation, say z of Z , which of the two distributions Z has.

When F and G are completely known, the solution to the problem is implicit in the Neyman-Pearson lemma. (1) The discrimination depends on the ratio $\frac{f(z)}{g(z)}$ where f and g are the respective density functions of F and G . The rule is as follows:

If $\frac{f(z)}{g(z)} > C$, decide in favor of F

If $\frac{f(z)}{g(z)} < C$, decide in favor of G

If $\frac{f(z)}{g(z)} = C$, the decision is arbitrary.

C is an appropriate positive constant chosen on the basis of consideration relating to the importance of the two possible errors:

- (i) $P_1 = P(Z \text{ is assigned to } G \mid Z \text{ came from } F)$
- (ii) $P_2 = P(Z \text{ is assigned to } F \mid Z \text{ came from } G).$

The two most widely advocated choices of C are:

(a) Take $C = 1$

(b) Choose C such that $P_1 = P_2$.

This procedure, known as the "likelihood ratio procedure" is known to have optimum properties with regard to control of the probability of misclassification.

When F and G are known except for the values of one or more parameters, the procedure used is much the same as that just described. Under the assumption that F and G are known except for one or more parameters and if we can assume that samples are available say:

$$\begin{array}{l} X_1, X_2, X_3, \dots, X_m \text{ from } F \\ Y_1, Y_2, Y_3, \dots, Y_n \text{ from } G \end{array}$$

we are able to estimate the unknown parameters, denoted collectively by θ . By some estimation procedure, we can estimate θ by $\hat{\theta}$ and assume that $F_{\hat{\theta}}$ and $G_{\hat{\theta}}$ are the correct distribution functions. The "likelihood ratio procedure" and the decision rules outlined above can now be applied.

If it is assumed that F and G are p -variate normal distributions having the same (unknown) covariance matrix and unknown expectation vectors, the linear discriminant function is a good example of this procedure. (2) The given samples are used to estimate the covariance matrices and the expectation vectors and the "likelihood ratio procedure" is used under the assumption that the estimated parameters are known to be correct. It is known that under the normal assumption for F and G and the homoscedastic assumption that the linear discriminant function is an optimal procedure.

Although this procedure seems reasonable when the parametric form of the distributions is correct or the assumed form is correct, there is concern about the validity of this procedure when the linear discriminant function is used with data not normal, or if normal, with unequal covariance matrices. In fact in the normal situation when the covariance matrices are not equal, a quadratic function can be shown to be optimal. There is a need then for a reasonable discrimination procedure whose validity does not require the knowledge implied by the normality assumption, the homoscedastic assumption or any assumption about the parametric form.

Several classes of nonparametric discrimination procedures were proposed in (3). These procedures were proven to have asymptotic optimum properties for large samples. In (4), some of these nonparametric procedures were investigated when the samples were small. These procedures were compared with the linear discriminant function where F and G were assumed normal with equal covariance matrices since under these assumptions the linear discriminant function is known to be optimal. A comparison was made by comparing the probabilities of misclassification when the linear discriminant function was used against the probabilities of misclassification when the nonparametric procedures were used. A survey of the procedures and results of (4) are given in



Section II of this paper.

In Section III of this paper, an investigation is made of the performance of one of the nonparametric discriminators discussed in (4) and of the performance of the linear discriminant function when F and G are not normal but, in fact, exponential with parameters λ and μ respectively. The exponential distribution was selected because of the role it plays in the field of life testing, and other applied problems. It is shown that for sample sizes of 1 and 2, that both the nonparametric discriminator and the linear discriminant function give very poor results for certain values of λ and μ .

Detailed conclusions and recommendations made on the basis of the results attained in Sections II and III are contained in Section IV of this paper.

Professors R. R. Read and J. R. Borsting, of the U. S. Naval Postgraduate School, have generously given their time to provide direction, encouragement and valuable advice to the author in the writing of this paper.

SECTION II

PERFORMANCE OF THE LINEAR DISCRIMINANT FUNCTION
AND A CLASS OF NONPARAMETRIC DISCRIMINATORS
WHEN THE TWO POPULATIONS BEING DISCRIMINATED
HAVE NORMAL DISTRIBUTIONS WITH
EQUAL COVARIANCE MATRICES

Let X_1, X_2, \dots, X_m be a sample from a p -variate distribution F and let Y_1, Y_2, \dots, Y_n be a sample from a p -variate distribution G . It is assumed further that the parametric forms of F and G are unknown. If z is an observation of a random variable Z known to be either distributed as F or G , how is it decided on the basis of z to which population Z belongs? Define a distance function (in p -dimensional space) which will permit a ranking of the $m+n$ observations according to their "nearness" to z . The idea of the discrimination procedures outlined in (3) is to assign Z to the population which has the most observations nearest to z . Specifically, choose an odd integer, k , and assume for simplicity that $m=n$, then Z is assigned to the distribution from which came the majority of the k nearest observations.

In (3), it was shown that several classes of these nonparametric discriminators have asymptotically optimum performance as $m \rightarrow \infty$ and $n \rightarrow \infty$ at the same rate. By optimum performance, it is meant that the probabilities of misclassification P_1 and P_2 , as defined in the introduction, tend to

the theoretical minimum values which they could have if F and G were completely known.

The asymptotic properties and the simplicity of applying the procedures of this class of nonparametric discriminators suggest that this type of procedures might be a reasonable alternative to the commonly applied linear discriminant function. However, to propose an alternative to the the linear discriminant function solely on the basis of asymptotic properties and ease of application would not be entirely reasonable. In particular, the small sample performance of such nonparametric discriminators needs investigation to ascertain how much discrimination power is lost when F and G are known to be normal with equal covariance matrices so that the linear discriminant function is appropriate. One way this investigation can be accomplished is by comparing the probabilities of misclassification when the linear discriminant function is used with the corresponding probabilities of misclassification when the nonparametric discriminators are used. Such an investigation was made in (4). The remainder of Section II is devoted to summarizing the procedures and results of (4).

It is first pointed out that the problem can be reduced considerably by considering linear transformations in the observation space. It is always possible by such transformations to insure that F and G will have the identity

covariance matrix. In other words, in the new space, the p transformed measurements are independent in each population and each measurement has a unit variance. It is also possible by such transformations to put the expectation vector of the F population at the origin and the expectation vector of the G population on the positive first axis. This allows complete specification of the transformed population by the two parameters p and λ where

$$\begin{aligned}\lambda &= E(\text{first coordinate of } Y) \\ &= \text{distance between the means of the} \\ &\quad \text{transformed populations.}\end{aligned}$$

In performing such linear transformations, P_1 and P_2 for the linear discriminant function are unchanged. The probabilities P_1 and P_2 for the nonparametric discriminators are likewise unchanged since such linear transformations map the totality of distance functions one-one into the totality in the new space.

It is assumed that the sizes of the samples taken from each population are equal, $m=n$. In the main, the distance function used is

$$\Delta(x, z) = \text{Max}_{i=1}^P |x_i - z_i|.$$

It should be pointed out that Δ is just one of a large class of distance functions, anyone of which could be used. This fact is mentioned since the probabilities P_1 and P_2 depend

very heavily on the distance function chosen. Most of the computations are made using $k=1$, that is, assign Z to the population F or G from which came the individual of the pooled samples which most closely resembles Z .

The first case considered is the univariate case, $p=1$. Using the rule of the "nearest neighbor"; that is, $k=1$, and the distance function $\Delta = |x-z|$, which corresponds to ordinary Euclidean distance in this case, the probabilities P_1 and P_2 are computed for various values of n and λ .

For $p=1$, the linear discriminant function is greatly reduced since no matrix computation enters. The arithmetic mean $\frac{\bar{X}+\bar{Y}}{2}$ of the sample means is computed and Z is assigned to that population whose sample mean lies on the side of $\frac{\bar{X}+\bar{Y}}{2}$ as does Z itself. The probabilities of misclassification are now readily computed.

From the symmetry of the problem, $P_1=P_2$ so it is sufficient to compute P_1 , thus, it is assumed that Z is distributed according to the F distribution. As was pointed out previously, linear transformations make it possible to put $E(X)=0$, $E(Y)=\lambda>0$ and $\sigma_X^2 = \sigma_Y^2 = 1$ with no loss of generality.

An error is committed by the linear discriminant function if and only if,

- (i) $Z > \frac{\bar{X}+\bar{Y}}{2}$ and $\bar{Y} > \bar{X}$
- (ii) $Z < \frac{\bar{X}+\bar{Y}}{2}$ and $\bar{Y} < \bar{X}$.

Define $U=\bar{Y}-\bar{X}$ and $V=\frac{\bar{X}+\bar{Y}}{2}-Z$. It is easily shown that U

very heavily on the distance function chosen. Most of the computations are made using $k=1$, that is, assign Z to the population F or G from which came the individual of the pooled samples which most closely resembles Z .

The first case considered is the univariate case, $p=1$. Using the rule of the "nearest neighbor"; that is, $k=1$, and the distance function $\Delta = |x-z|$, which corresponds to ordinary Euclidean distance in this case, the probabilities P_1 and P_2 are computed for various values of n and λ .

For $p=1$, the linear discriminant function is greatly reduced since no matrix computation enters. The arithmetic mean $\frac{\bar{X}+\bar{Y}}{2}$ of the sample means is computed and Z is assigned to that population whose sample mean lies on the side of $\frac{\bar{X}+\bar{Y}}{2}$ as does Z itself. The probabilities of misclassification are now readily computed.

From the symmetry of the problem, $P_1=P_2$ so it is sufficient to compute P_1 , thus, it is assumed that Z is distributed according to the F distribution. As was pointed out previously, linear transformations make it possible to put $E(X)=0$, $E(Y)=\lambda>0$ and $\sigma_X^2 = \sigma_Y^2 = 1$ with no loss of generality.

An error is committed by the linear discriminant function if and only if,

- (i) $Z > \frac{\bar{X}+\bar{Y}}{2}$ and $\bar{Y} > \bar{X}$
- (ii) $Z < \frac{\bar{X}+\bar{Y}}{2}$ and $\bar{Y} < \bar{X}$.

Define $U=\bar{Y}-\bar{X}$ and $V=\bar{X}+\bar{Y}-2Z$. It is easily shown that U

and V are independent normal random variables with $E(U)=\lambda$, $\sigma_U^2 = 2/n$, $E(V)=\lambda$, $\sigma_V^2 = 4 + 2/n$. In terms of the variables U and V , an error is committed by the linear discriminant function if and only if $UV < 0$. Thus it follows for the linear discriminant function when $p=1$, that,

$$P_1 = P_2 = \left[1 - \phi\left(-\frac{\sqrt{n}\lambda}{\sqrt{2}}\right) \right] \phi\left(-\frac{\sqrt{n}\lambda}{\sqrt{2+4n}}\right) + \phi\left(-\frac{\sqrt{n}\lambda}{\sqrt{2}}\right) \left[1 - \phi\left(-\frac{\sqrt{n}\lambda}{\sqrt{2+4n}}\right) \right],$$

where

$$\phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

Since $\lim_{n \rightarrow \infty} P = \phi(-\frac{\lambda}{2})$, it is observed that the maximum probability of misclassification is .5. The values of $P_1 = P_2$ for various values of n and λ are given in Table 1. Figures 1 and 2 give these results graphically. All Tables and Figures in Section II have been reproduced from (4).

We consider now the nonparametric discriminator using the "rule of the nearest neighbor," $k=1$, which consists of assigning Z to that population from which came the sample individual nearest to z . Suppose that $Z=z$. Let $P_1(z)$ denote the conditional probability that the nearest of the $2n$ sample observations to z is a y , given $Z=z$. Then,

$$P_1 = E(P_1(Z)) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} P_1(z) dz \quad (1)$$

TABLE 1
PROBABILITY OF ERROR, LINEAR DISCRIMINANT FUNCTION,
UNIVARIATE NORMAL DISTRIBUTIONS

n	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
1	.4175	.2532	.1235
2	.3821	.1999	.0910
3	.3611	.1819	.0826
4	.3472	.1744	.0787
5	.3376	.1707	.0763
10	.3175	.1646	.0716
20	.3110	.1616	.0692
50	.3094	.1599	.0678
∞	.3085	.1587	.0668

n = size of sample taken from each population

λ = distance between the means of the two populations

Probability of error = $P(Z \text{ is assigned to } G \mid Z \text{ came from } F)$

= $P(Z \text{ is assigned to } F \mid Z \text{ came from } G)$

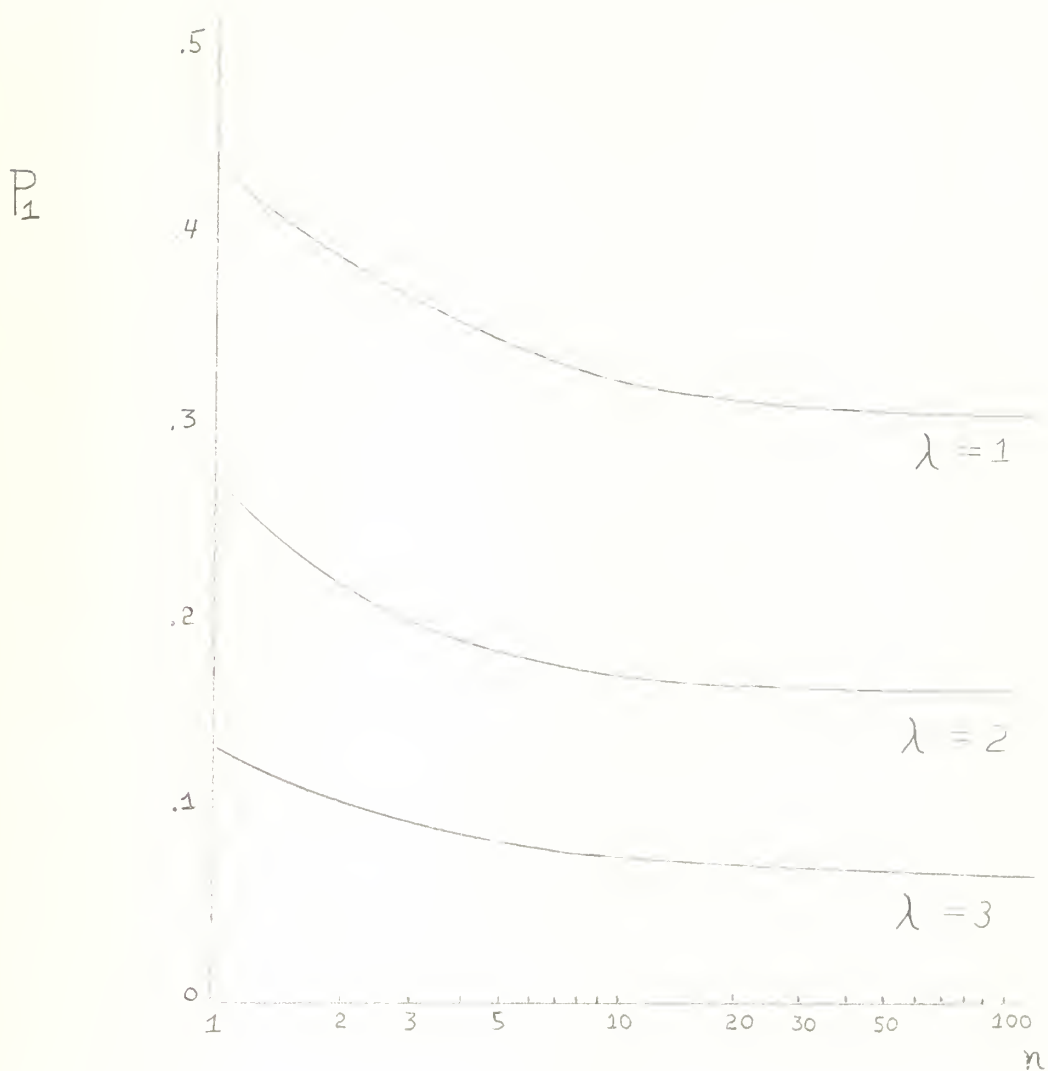


FIGURE 1

Probability of error P_1 of the linear discriminant function for two univariate normal distributions with distance between means $= \lambda$.

n = size of sample from each population.

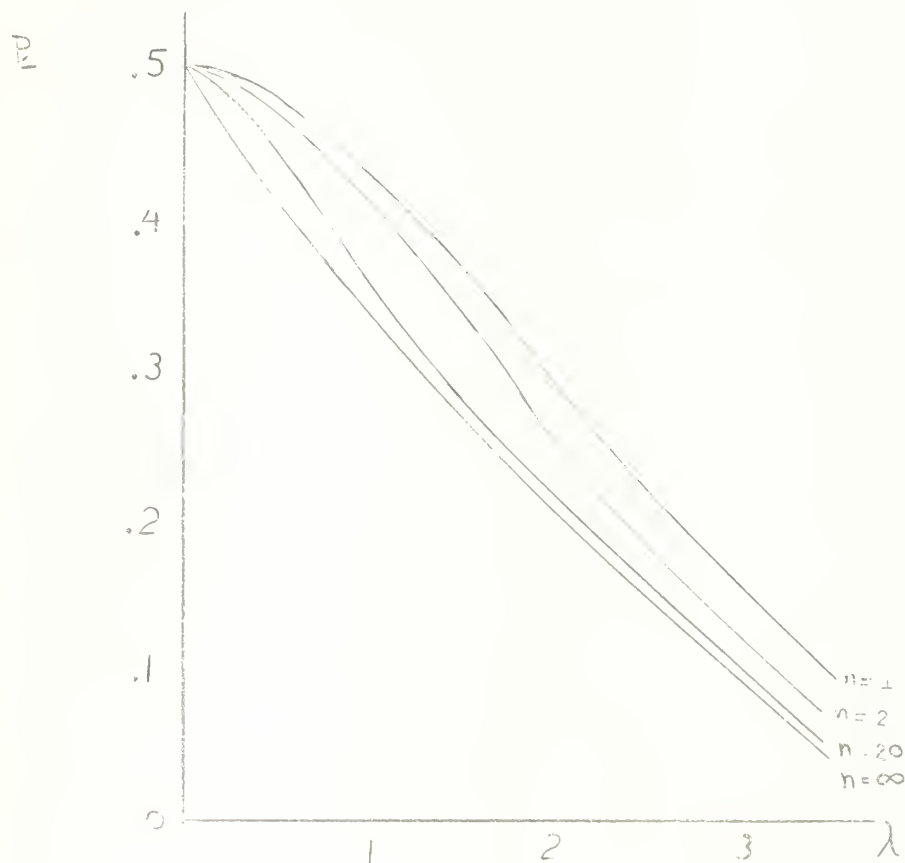


FIGURE 2

Probability of error P_1 of the linear discriminant function for two univariate normal distributions with distance between the means $= \lambda$, plotted as a function of λ .

n = size of sample from each population.

It remains then to calculate $P_1(z)$. Define

$$\begin{aligned} H_z(\delta) &= P(|X - z| < \delta) \quad \delta > 0 \\ &= P(z - \delta < X < z + \delta) \\ &= \Phi(z + \delta) - \Phi(z - \delta), \end{aligned}$$

and

$$\begin{aligned} K_z(\delta) &= P(|Y - z| < \delta) \\ &= P(z - \lambda - \delta < Y - \lambda < z - \lambda + \delta) \\ &= \Phi(z - \lambda + \delta) - \Phi(z - \lambda - \delta). \end{aligned}$$

The event, "the nearest sample value to z is a y " can be classified into the n exclusive equiprobable events, "the nearest sample value to z is y_i , $i=1, 2, \dots, n$. Since the nearest y to z will necessarily be the minimum y , it is necessary to compute the probability density function for the minimum of $|Y_1 - z|$, $|Y_2 - z|$, \dots , $|Y_n - z|$. Since the $|Y_i - z|$, $i = 1, 2, \dots, n$, are independent identically distributed random variables, this density function is easily shown to be,

$$n(1 - K_z(\delta))^{n-1} dK_z(\delta).$$

$P_1(z)$ is then computed by the following formula:

$$P_1(z) = n \int_0^\infty (1 - H_z(\delta))^n (1 - K_z(\delta))^{n-1} dK_z(\delta) \quad (2).$$

Formulae (1) and (2) form the basis for all the computations for the "nearest neighbor rule" no matter what the value of p if for $p > 1$ one replaces $P(|X - z| < \delta)$ by P (the distance of X from $z < \delta$) and similarly $P(|Y - z| < \delta)$ by P (the distance of Y from $z < \delta$). Of course the specific evaluations depend upon the distance function used.

Except for the case $p=1$, $n=1$, in which case P_1 and P_2 are the same for the linear discriminant function and the nonparametric discriminator, the bulk of the computations for the nonparametric discriminator were carried out by straightforward numerical integration. These computations are given in Table 2. These computations are quite heavy, especially for the case $p=2$. Therefore, a search for an approximation formula for the computation of $P_1(z)$ was instituted. One approximation formula was found which gave very good results. A discussion of this approximation formula is given in (4), P_1 as computed using the approximation formula for $P_1(z)$ is tabled in Table 2-A. One very interesting result which was obtained using the approximation formula for $P_1(z)$ was that for large n ,

$$P_1 \cong E \left[\frac{g(z)}{f(z) + g(z)} \right] = \int_{-\infty}^{\infty} \frac{f(z)g(z)}{f(z) + g(z)} dz.$$

An application of Schwartz's inequality shows the latter integral to be at most 0.5. It is thus possible to assert that, whatever be the populations being discriminated, the "rule of the nearest neighbor" will have in the limit as $m = n \rightarrow \infty$ equal probabilities of error at most 0.5.

To compare the figures of Tables 1, 2, and 2-A, the values of $P_1 = P_2$ for paired values of λ are plotted against n in Figure 3. In Figure 4, the same values are plotted against λ for selected values of n .

TABLE 2
PROBABILITY OF ERROR, NONPARAMETRIC DISCRIMINATOR
WITH $k=1$, UNIVARIATE NORMAL DISTRIBUTION

n	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
1	.4175	.2532	.1235
2	.4086	.2364	.1084
3	.4052	.2307	.1036
4	.4032	.2280	.1014

TABLE 2-A
APPROXIMATE PROBABILITY OF ERROR, NONPARAMETRIC
DISCRIMINATOR WITH $k=1$, UNIVARIATE NORMAL DISTRIBUTION

n	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
4	.403	.226	.102
5	.401	.225	.100
10	.399	.223	.098
20	.398	.224	.098
50	.398	.225	.098
∞	.398	.225	.098

n = size of sample from each population

λ = distance between the means of the two populations

Probability of error = $P(Z \text{ is assigned to } G \mid Z \text{ came from } F)$

= $P(Z \text{ is assigned to } F \mid Z \text{ came from } G)$

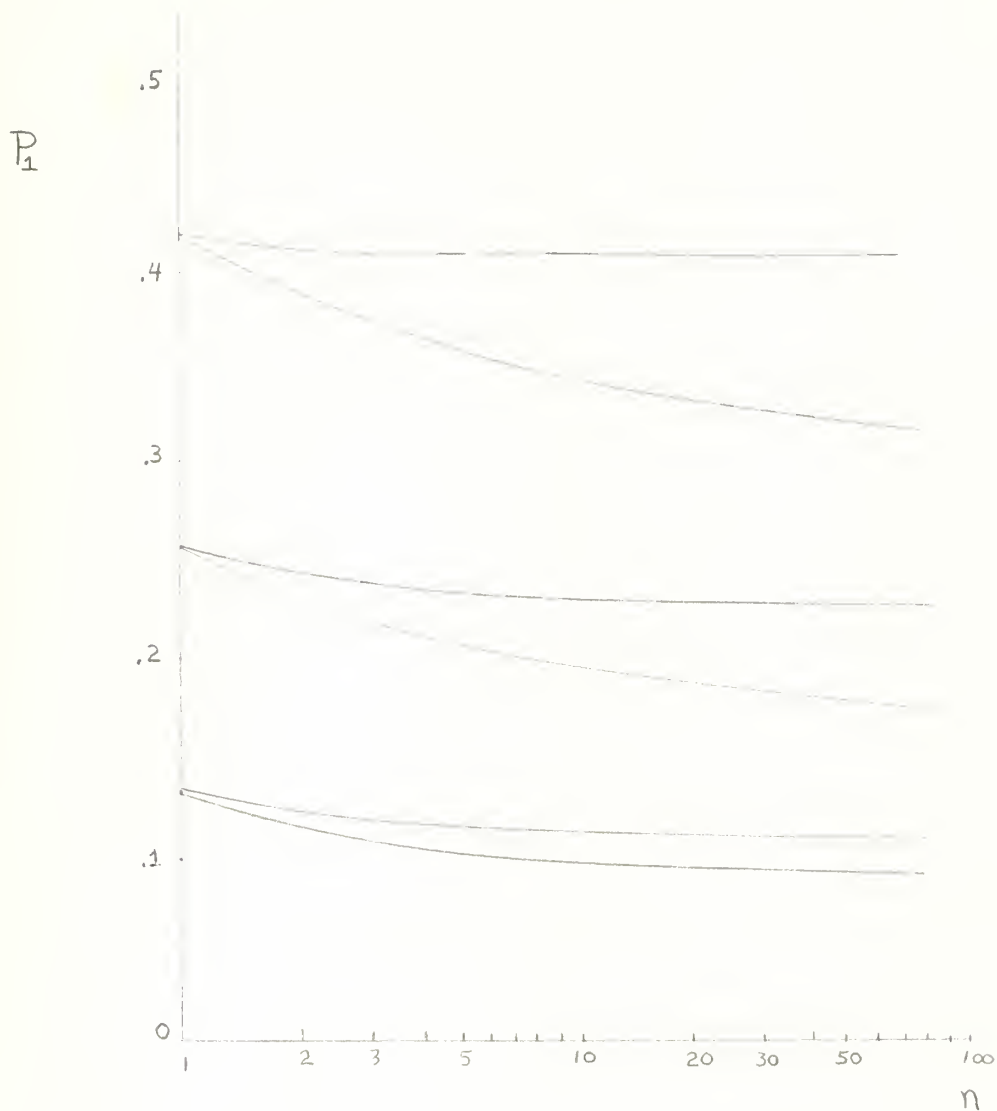


FIGURE 3

Comparison of the probability of error P_1 as a function of n for the linear discriminant function and the nonparametric discriminator, distance function Δ , $k=1$, for two normal univariate populations with distance between means = λ .
 n = size of sample from each population

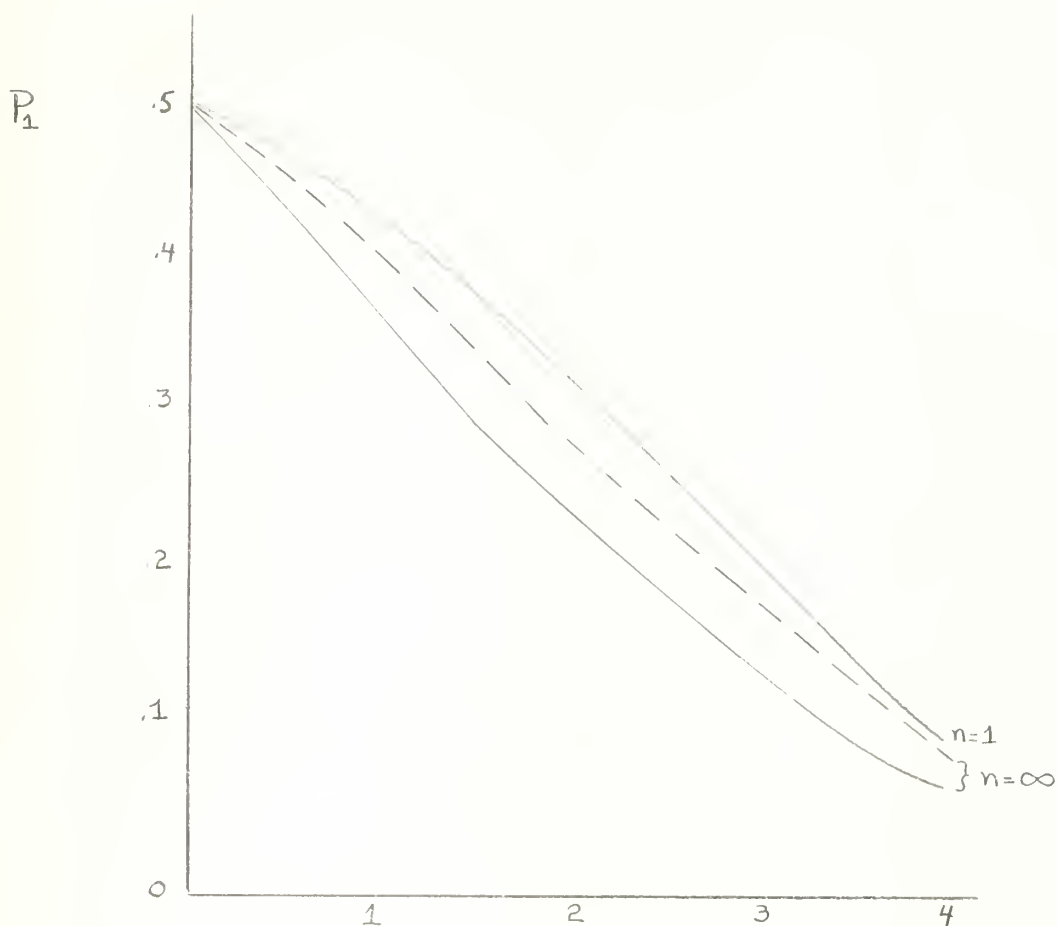


FIGURE 4

Comparison of the probability of error P_1 as a function of λ , the distance between the means, for the linear discriminant function and the nonparametric discriminator, distance function = Δ , $k=1$, for two normal univariate populations

n = size of sample from each population

$n = 1$ is identical for both

--- indicates the nonparametric procedure

Not discussed in this paper, but investigated to a very limited extent in (4) are the following cases:

- (i) the nonparametric discriminator using Δ as a distance function with $k \geq 3$ for the univariate and bivariate normal distributions
- (ii) the nonparametric discriminator using Δ as a distance function $k = 1$, $n = 1$ for $p \geq 2$
- (iii) the effect of distance functions other than Δ on the probabilities of misclassification for bivariate normal distribution

Although the investigation of the above cases was extremely limited due to the laborious computations, the results that were obtained indicated that the nonparametric discrimination procedure gave "reasonable" error probabilities in both cases (i) and (ii). In the bivariate normal distribution, different distance functions produced vastly different error probabilities in some situations.

SECTION III

PERFORMANCE OF THE LINEAR DISCRIMINANT FUNCTION AND A CLASS OF NONPARAMETRIC DISCRIMINATORS WHEN F AND G ARE EXPONENTIALLY DISTRIBUTED

In this section, a limited investigation of the linear discriminant function and the nonparametric discriminator using Δ as a distance function and using "the rule of the nearest neighbor," $k=1$, is made when F and G are not normally distributed; but in fact, exponentially distributed with parameters λ and μ respectively. The performance of both the linear discriminant function and the nonparametric discriminator will be investigated again by computing the probabilities of misclassification. Under the assumption that F and G are exponentially distributed, it will be shown that the linear discriminant function and the nonparametric discriminator using Δ as a distance function and "the rule of the nearest neighbor" can give high probabilities of misclassification.

Throughout the remainder of the section, it will be assumed that $m = n$ and that F and G are exponentially distributed with parameters λ and μ respectively. Because of the heavy computations involved in computing the probabilities of misclassification;

- (i) $P_1 = P(\text{assigning } Z \text{ to } G \mid Z \text{ came from } F)$
- (ii) $P_2 = P(\text{assigning } Z \text{ to } F \mid Z \text{ came from } G)$

the only cases investigated will be for $p=1$ and $n=1,2$.

P_1 and P_2 will first be computed for the linear discriminant function. The procedure here is precisely that which was used in Section II for $p = 1$. One simply computed the arithmetic mean $\frac{\bar{X} + \bar{Y}}{2}$ of the sample means and assigns Z to that population whose sample mean lies on the side of $\frac{\bar{X} + \bar{Y}}{2}$ as does z itself. While $P_1 \neq P_2$, it is only necessary to compute P_1 since P_2 can readily be computed from P_1 by interchanging λ and μ .

Proceeding as in Section II, define the new variables $U = \bar{Y} - \bar{X}$ and $V = \bar{X} + \bar{Y} - 2Z$. If U and V are to be independent, it is necessary that the covariance of U and V be zero. Computing the covariance of U and V we have:

$$\text{Cov}(U, V) = \frac{1}{n} \left(\frac{1}{\lambda} - \frac{1}{\mu^2} \right) \neq 0 \text{ except for } \lambda = \mu.$$

Since discrimination is not possible for $\lambda = \mu$, the $\text{Cov}(U, V)$ will not be zero and in general U and V will not be independent. As before, an error is committed by linear discriminant function if and only if;

$$\begin{aligned} \text{(i)} \quad Z &> \frac{\bar{X} + \bar{Y}}{2} \quad \text{and} \quad \bar{Y} > \bar{X} \\ \text{(ii)} \quad Z &< \frac{\bar{X} + \bar{Y}}{2} \quad \text{and} \quad \bar{Y} < \bar{X}. \end{aligned}$$

In terms of the variables U and V , an error is committed if and only if $UV < 0$, and therefore,

$$P_1 = P(UV < 0).$$

Since U and V are not in general independent, the probability that $UV < 0$ is not easily computed. It is necessary to compute the joint density function for U and V and integrate over the region where $UV < 0$. The joint density function of U and V was computed but because of the complex nature of this function, it was considered easier to compute P_1 directly. By (i) and (ii) and the definition of P_1 it follows that,

$$P_1 = P \left(Z > \frac{\bar{X} + \bar{Y}}{2}, \bar{Y} > \bar{X} \right) + P \left(Z < \frac{\bar{X} + \bar{Y}}{2}, \bar{Y} < \bar{X} \right).$$

Let $T = n\bar{Y}$ and $S = n\bar{X}$ and thus,

f_T is the gamma density function with parameters n and μ

f_S is the gamma density function with parameters n and λ .

Since T , S , and Z are independent random variables,

$$P_1 = \int_0^\infty \int_0^\infty \int_{\frac{T+S}{2n}}^\infty f_Z(z) f_T(t) f_S(s) dz dt ds \\ + \int_0^\infty \int_0^s \int_0^{\frac{T+S}{2n}} f_Z(z) f_T(t) f_S(s) dz dt ds.$$

P_1 can now be computed by direct numerical integration. For $n=1$, P_1 as a function of λ and μ is,

$$P_1(\lambda, \mu) = \frac{\mu(10\lambda^2 + 2\mu^2 + 15\lambda\mu)}{3(\mu + 2\lambda)(\mu + \lambda)(2\mu + \lambda)}.$$

By interchanging λ and μ , P_2 is,

$$P_2(\lambda, \mu) = P_1(\mu, \lambda)$$

Recognizing that the numerator and denominator in the expressions for P_1 and P_2 are homogeneous of degree 3 in λ and μ , P_1 and P_2 can be expressed in terms of a single parameter c by setting $\lambda = c\mu$. Making this substitution in the expressions for P_1 and P_2 we have,

$$P_1(c) = \frac{(10c^2 + 15c + 2)}{3(1 + 2c)(1 + c)(2 + c)}$$

$$P_2(c) = P_1\left(\frac{1}{c}\right)$$

For $n=1$, P_1 and P_2 for the linear discriminate function are the same as P_1 and P_2 for the nonparametric discriminator using Δ as a distance function and "the nearest neighbor rule, $k=1$."

For $n=2$, the substitution $\lambda = c\mu$ is again appropriate and P_1 and P_2 for $n=2$ are as follows,

$$P_1(c) = \frac{128c^2(2c + 3)}{(c + 4)^2 (3c + 2)^3} + \frac{(3c + 1)}{(c + 1)^3} - \frac{128(4c + 1)}{25(3c + 2)^3}$$

$$P_2(c) = P_1\left(\frac{1}{c}\right)$$

Values of P_1 and P_2 for the linear discriminant function for $n=1$ and 2 are tabled for various values of c in Table 3.

P_1 and P_2 are next computed for the nonparametric discriminator for the case $n=2$. The procedure used is exactly the procedure used in Section II. The substitution $\lambda = c\mu$, is once more appropriate. P_1 and P_2 in terms of a single parameter c are as follows:

$$\begin{aligned}
P_1(c) = & \frac{(30c^2 - 38c - 112)}{15(c+2)(c-1)} + \frac{(32 + 24c - 56c^2 - 12c^3)}{3(3c+2)(c^2-1)} \\
& + \frac{16c^2}{(c^2-1)(2c+1)} + \frac{4}{(5c+2)} + \frac{(112 - 52c - 30c^2)}{15(c+4)(c-1)} \\
& + \frac{4(3c+8)}{3(3c+4)} - \frac{1}{(c+1)} \quad \text{for } c \neq 2
\end{aligned}$$

$$\begin{aligned}
P_1(c) = & \frac{(30c^2 - 38c - 112)}{15(c+2)(c-1)} + \frac{(32 + 24c - 56c^2 - 12c^3)}{3(3c+2)(c^2-1)} \\
& + \frac{(2c^3 + 16c^2 - 2c)}{(2c+1)(c^2-1)} + \frac{4}{(5c+2)} + \frac{(24 - 4c - 10c^2)}{5(c+4)(c-1)} \\
& - \frac{(3c+1)}{(c+1)^2} + \frac{4(3c+2)}{3(3c+4)} \quad \text{for } c = 2
\end{aligned}$$

$$P_2(c) = P_1\left(\frac{1}{c}\right).$$

Values of P_1 and P_2 for various values of c for the non-parametric discriminator with $n=2$ are given in Table 3.

It is observed in Table 3, that P_1 and P_2 exceed 0.5 for numerous values of c . Because of this observation, an investigation was made to determine the values of c for which P_1 and P_2 exceed 0.5. Figure 5 displays graphically the regions in the λ, μ plane where P_1 and P_2 are greater than 0.5.

Figure 5, points out only too well that great caution should be used when applying the linear discriminant in situations when the populations are other than normal.

TABLE 3
PROBABILITIES OF ERROR, UNIVARIATE
EXPONENTIAL DISTRIBUTIONS

	c	1	2	3	4	5	10
Linear Discriminant	P_1	.5000	.4000	.3262	.2741	.2360	.1385
Function*, n=1	P_2	.5000	.5333	.5214	.5037	.4870	.4329
Linear Discriminant	P_1	.5000	.3736	.2652	.2009	.1567	.0627
Function, n=2	P_2	.5000	.5299	.5041	.4782	.4577	.4056
Nonparametric Discriminator, n=2	P_1	.5000	.4222	.3559	.3066	.2692	.1675
	P_2	.5000	.5003	.4666	.4295	.3706	.3281

c is a parameter such that $\lambda = c\mu$

λ is the parameter of the F population

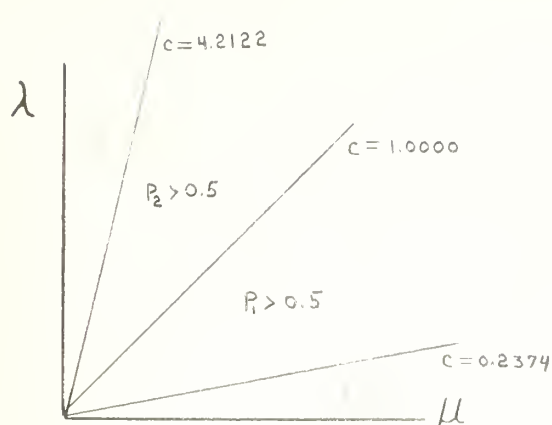
μ is the parameter of the G population

$P_1 = P(\text{assigning } Z \text{ to } G \mid Z \text{ came from } F)$

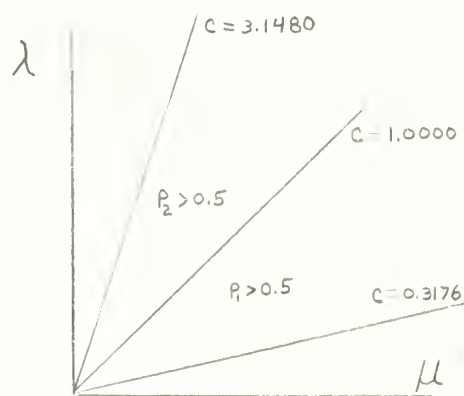
$P_2 = P(\text{assigning } Z \text{ to } F \mid Z \text{ came from } G)$

n = sample size

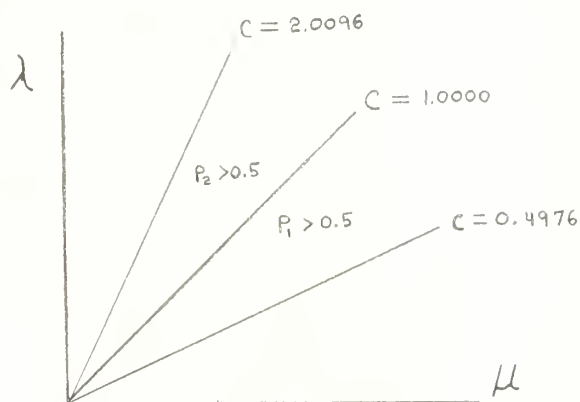
*For n=1, the probabilities of error P_1 and P_2 for the linear discriminant function are equal to the corresponding probabilities of error P_1 and P_2 for the nonparametric discriminator.



Linear discriminant
function*, $n = 1$



Linear discriminant
function, $n = 2$



Nonparametric discriminator
 $n = 2$

FIGURE 5

Values of λ and μ for which P_1 and P_2 exceed 0.5

c = parameter such that $\lambda = c \mu$

λ = parameter of F distribution

μ = parameter of G distribution

$P_1 = P(Z \text{ is assigned to } G \mid Z \text{ came from } F)$

$P_2 = P(Z \text{ is assigned to } F \mid Z \text{ came from } G)$

n = sample size

*Linear discriminant function is equivalent to the nonparametric discriminator for $n = 1$.

SECTION IV

SUMMARY AND CONCLUSIONS

In any discrimination problem one has a choice between using parametric or nonparametric procedures. This choice in general will depend upon three factors:

- (i) the strength of the users belief in his parametric model.
- (ii) the loss that would be suffered by using the nonparametric rule if in fact the parametric form is correct.
- (iii) the loss that would be suffered by using the parametric rule if the actual densities depart from the parametric form assumed.

For the two population discrimination problem, Section II of this paper concerned itself with (ii). In Section II, it was assumed that the two populations being discriminated were normal with equal covariance matrices. For the univariate case, the parametric procedure used was the well known linear discriminant function which is known to be optimal in this situation. The nonparametric procedure used was the rule whereby a random variable was classified as belonging to the population which had the nearest observation to an observed value of the random variable being classified. A comparison of these two procedures was made by computing and comparing the probabilities of misclassification.

Also for the two population discrimination problem, an investigation of the linear discriminant function and the same nonparametric procedure was carried out when the two populations were not normal but exponential. Again the investigation was made by computing the probabilities of misclassification for both procedures. This investigation was made in Section III of this paper. Because of the lengthy computations involved in computing the probabilities of error for both of these procedures, the only cases considered were the univariate case for sample sizes of 1 and 2. It was shown that for the two cases investigated, sample sizes of 1 and 2, that both the procedures could give poor results depending on the parameters of the distributions.

In conclusion, it seems reasonable that if the populations to be discriminated are well known, and have been investigated to be such that the normal distribution gives a good fit and that the variance and correlation do not change much when the means are changed, and if the classification to be made warrants the labor of matrix inversion, then the linear discriminant function should be used. However, if the populations are either not well known; or are known not to be approximately normal or to have very different covariance matrices; or if the discrimination is such that small decreases in probability of error are not worth extensive computations, then a nonparametric procedure seems

to be advisable. Which nonparametric procedure is a matter of choice for the user.

Recommendations to be made on the basis of this paper are:

- (i) tabulate the probabilities of error for the linear discriminant function in representative situations for the case where the populations being discriminated are multivariate normal with equal covariance matrices.
- (ii) further investigation (for larger sample sizes) of the linear discriminant function in the case where the populations being discriminated are exponential because of the importance of the exponential distribution in the field of life testing and other applied problems.
- (iii) investigation as to the effect of other distance functions for the nonparametric discriminator discussed in this paper in the case when the populations being discriminated are exponential or some other class of distributions.

BIBLIOGRAPHY

1. J. Neyman and E. S. Pearson, "Contributions to the Theory of Testing Statistical Hypotheses," Stat. Res. Memoirs, Vol. I (1936), pp. 1-37.
2. R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, Vol. 7 (1936), pp. 179-188.
3. Evelyn Fix and J. L. Hodges, Jr., "Nonparametric Discrimination. I. Consistency Properties," Technical Report prepared under contract between University of California and School of Aviation Medicine, Randolph Field, Texas, 22 pp.
4. Evelyn Fix and J. L. Hodges Jr., "Nonparametric Discrimination: Small Sample Performance," Technical Report prepared under contract between University of California and School of Aviation Medicine, Randolph Field, Texas.

thesE144

Discrimination procedures, small sample



3 2768 001 90287 7

DUDLEY KNOX LIBRARY